



# Réutilisation d'un corpus pour une nouvelle analyse des réseaux sociaux grâce à l'adaptation de l'outil KSV

Christophe Reffay, Christopher Teplovs

## ► To cite this version:

Christophe Reffay, Christopher Teplovs. Réutilisation d'un corpus pour une nouvelle analyse des réseaux sociaux grâce à l'adaptation de l'outil KSV. Environnements Informatiques pour l'Apprentissage Humain, May 2011, Mons, Belgique. pp.139-150. edutice-00596729v2

**HAL Id: edutice-00596729**

**<https://edutice.archives-ouvertes.fr/edutice-00596729v2>**

Submitted on 31 May 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Réutilisation d'un corpus pour une nouvelle analyse des réseaux sociaux grâce à l'adaptation de l'outil KSV

## Conférence EIAH 2011 (Environnements Informatiques pour l'Apprentissage Humain)

**Christophe Reffay\*, Christopher Teplovs\*\***

\* UMR Science Technique Education Formation,  
INRP - ENS Cachan - UniverSud  
61, avenue du Président Wilson  
F-94235 Cachan cedex  
christophe.reffay@inrp.fr

\*\* Department of Curriculum, Teaching and Learning  
Ontario Institute for Studies in Education, University of Toronto  
252 Bloor Street West,  
Toronto, Canada M5S 1V6  
chris.teplovs@utoronto.ca

*RÉSUMÉ. Dans le champ de l'analyse des interactions en situation d'apprentissage collaboratif en ligne, nous revisitons un corpus connu et accessible pour en reprendre l'analyse avec un outil qui offre de nouvelles dimensions d'exploration : l'intensité et le temps. Le cadre d'analyse est ici celui des réseaux sociaux. Cette nouvelle exploration des données nous permet de proposer de nouveaux indicateurs pour comparer la cohésion des groupes au regard de l'intensité de communication. Cette rencontre entre un corpus et un outil a aussi permis d'étendre les capacités de traitement et de visualisation de l'outil KSV, qui lui offrent de nouvelles perspectives.*

*MOTS-CLÉS : Réseaux sociaux, Forum, Collaboration, Cohésion, SNA, CMC.*

## 1. Introduction

C'est avant tout la qualité de l'outil [KSV] (Knowledge Space Visualizer) [TEPLOVS 10], qui a suscité l'envie de l'appliquer à l'analyse des réseaux sociaux dans des forums de discussions alors qu'il était conçu pour révéler les relations sémantiques entre les contributions dans l'environnement Knowledge Forum [SCARDAMALIA & BEREITER 03]. Il intègre des curseurs permettant de faire varier les seuils aisément, et peut être adapté. Une analyse des réseaux sociaux déjà appliquée aux interactions dans les forums de Simuligne a été présentée dans [REFFAY & CHANIER 03a]. Ce premier travail, réalisé avec les outils [Ucinet 6] et [Graphviz], avait permis de construire les sociogrammes à partir desquels une mesure de la cohésion avait été proposée.

Pour mieux montrer le bénéfice que l'on peut espérer en réalisant l'analyse avec l'outil KSV, nous avons choisi de reprendre l'ensemble de données Simuligne, utilisé dans [REFFAY & CHANIER 03a]. Cette réutilisation a été facilitée par le travail de partage de corpus d'apprentissage réalisé dans le projet [Mulce]. Le corpus Simuligne est en effet librement accessible depuis 2009 sur la plateforme [Mulce]. D'un point de vue méthodologique, ce travail doit donc être considéré comme une nouvelle analyse appliquée à un ensemble de données déjà présenté. En procédant de la sorte, plutôt qu'en analysant un nouvel ensemble de données, nous pourrions garantir que les résultats des nouvelles analyses sont obtenus grâce au potentiel des nouveaux outils et ne sont pas des phénomènes qui auraient pu être nouveaux dans un nouveau jeu de données.

Au-delà du bénéfice pour l'analyse, la rencontre entre le corpus Simuligne et l'outil KSV a aussi profité au développement de l'outil. Pour réussir son adaptation, KSV a dû prendre en compte de nouvelles structures de données et modifier en partie son interface. Ce faisant, la version qui en résulte est indépendante du Knowledge Forum et ouvre pour KSV de nouveaux horizons d'application.

Notre communication est structurée en 5 parties. À la suite de cette introduction, la deuxième rappelle les acquis antérieurs concernant le corpus Simuligne et les résultats connus de l'analyse des réseaux sociaux appliquée aux interactions en formation en ligne. La troisième présente l'outil KSV dans sa forme originelle ainsi que les adaptations qui ont été réalisées pour ce travail. Nous pourrions alors mieux rendre compte en quatrième partie, de l'exploration multi dimensionnelle du corpus Simuligne avec la nouvelle version de KSV et présenter d'une part les avancées dans le champ de l'analyse des réseaux sociaux et d'autre part le bénéfice de cette expérience pour KSV. La conclusion permettra de dresser le bilan de ce travail et les perspectives qu'il offre à notre communauté de recherche.

## 2. Contexte et acquis antérieurs

Afin de mieux planter le décor, nous commencerons ici par présenter le corpus Simuligne et les conditions de sa réutilisation. Nous pourrions alors faire une brève introduction du cadre d'analyse des réseaux sociaux avant de présenter un état de l'art sur leur application à l'analyse des interactions en ligne essentiellement issu de la communauté CSCL (*Computer Supported Collaborative Learning*). Nous pourrions alors rappeler les principaux résultats de la première analyse [REFFAY & CHANIER 03a].

## **2.1. Réutilisation du Corpus Simuligne**

Simuligne est une situation d'apprentissage collaboratif en ligne du français langue étrangère qui a été réalisée en 2001. Elle met en œuvre la méthode de la simulation globale [YAICHE 96], habituellement utilisée comme méthode intensive en présentiel pour l'apprentissage des langues. Elle a été adaptée à une formation en ligne et jouée par 4 groupes d'apprenants 100% à distance.

Cette situation compte 40 apprenants (anglais en formation continue inscrits l'Open University en Angleterre), 10 natifs (étudiants francophones en master FLE à Besançon), servant de référents pour la langue et la culture française et assistant les 4 tuteurs (enseignants de FLE de l'Open University) eux-mêmes coordonnés par un responsable pédagogique. Les apprenants et natifs ont été répartis en 4 groupes de base (nommés : Aquitania, Gallia, Lugdunensis et Narbonensis) qui ont joué la simulation globale en parallèle et sans que les apprenants puissent échanger entre groupes. Un espace Formateurs, servant à la coordination des groupes et à la communication entre coordinateur et tuteurs était aussi accessible aux 10 natifs. Indépendamment de la simulation globale, un espace Monde a servi à dérouler l'activité « Interculture », inspirée de « Cultura » [FURSTENBERG et al. 01] où chaque participant (français ou anglais) pouvait réagir puis s'exprimer librement (en français ou en anglais) sur des mots, des phrases et des situations choisies sur des thèmes où ces deux cultures divergent, voire s'opposent. Si, dans la simulation globale, chaque groupe de base était clairement dirigé par son tuteur, personne n'était désigné en particulier pour encadrer l'activité « Interculture » dans le groupe Monde.

Trois des quatre groupes de base ont achevé la simulation globale, ce qui constitue un taux de réussite élevé en formation à distance. Le 31 mai, le groupe Lugdunensis a été déclaré défaillant et deux de ses membres (L14 et L19) ont été transférés dans le groupe Aquitania. L'ensemble de la formation Simuligne a généré un volume de texte relativement important (1 531 086 caractères) dans les outils de communication : forum (57%), courriel (27%) et chat (15%) dans les 4 groupes de base, le groupe Formateur et le groupe Monde. La prépondérance des forums dans cette formation nous a conduit à concentrer nos analyses sur les traces de cet outil de communication. Les analyses présentées ici ne prennent donc pas en compte les interactions qui ont eu lieu par courriel ou par chat.

Le corpus Simuligne, son contexte (scénario pédagogique, protocole de recherche) et l'ensemble des interactions (au format XML), a été anonymisé et déposé sur la plateforme [Mulce] depuis 2009 où il a reçu l'identifiant « mce.simu.all.all ». Ces données, ainsi structurées, ont pu être facilement transformées et réinjectées dans KSV.

## **2.2. Les réseaux sociaux : introduction**

Le cadre des réseaux sociaux [DEGENNES & FORSÉ 94], ou analyse structurale, s'oppose en sociologie à l'analyse statistique des catégories. Il est issu de la théorie des graphes, de la sociométrie, de la dynamique des groupes et de leur intersection avec l'anthropologie [SCOTT 00]. Dans ce cadre, on considère un type de relation et un groupe d'individus (non nécessairement fixé a priori). On représente ensuite le graphe constitué de nœuds (les individus finalement concernés) et d'arcs (les relations répertoriées). Dans ce cadre, un tel graphe est appelé sociogramme. Dès lors, on peut utiliser la panoplie des

algorithmes et des représentations qui révèlent les propriétés de la structure du sociogramme ainsi que certaines positions caractéristiques des individus dans le groupe. Selon le contexte et la définition des relations, ces propriétés mises en évidence peuvent être interprétées pour faire un diagnostic sur l'état des relations dans le groupe et la position que les individus y occupent. Les propriétés les plus utilisées sont l'équivalence, la cohésion, les diverses centralités (de degré, de proximité ou d'intermédiarité). Des travaux de sociologies ont aussi montré *la force des liens faibles* ou l'intérêt des *trous structuraux* et mettent en évidence le *capital social* des individus situés à l'articulation des différentes communautés ou encore le pouvoir des individus centraux.

Dans ce travail, nous nous intéressons plus spécialement à la cohésion d'un groupe et nous utilisons le cadre des réseaux sociaux pour tenter de la mesurer. Cette mesure étant basée sur la notion de cliques, nous en rappelons les définitions :

- Clique simple (dans un graphe non orienté, non valué) : sous-ensemble (maximum) de nœuds où chacun d'eux est connecté à tous les autres (i.e : un sous-graphe complet).
- k-clique (dans un graphe non orienté, non valué) : sous-ensemble de k nœuds où chacun d'eux est connecté à tous les autres (i.e : un sous-graphe complet à k noeuds).
- Clique de niveau s (dans un graphe non orienté, valué) : sous-ensemble (maximum) de nœuds tous connectés par des arcs de valeur au moins égale à s.
- k-clique de niveau s (graphe non orienté, valué) : sous-ensemble de k nœuds tous connectés par des arcs de valeur au moins égale à s.

Par définition, la cohésion d'un groupe est d'autant plus forte dans un groupe, que le nombre et l'intensité des relations sont grands. Ainsi, nous montrerons comment KSV nous permet de visualiser les cliques (de différentes tailles k) pour différentes valeurs d'intensité.

### **2.3. Les réseaux sociaux et l'analyse des interactions en ligne**

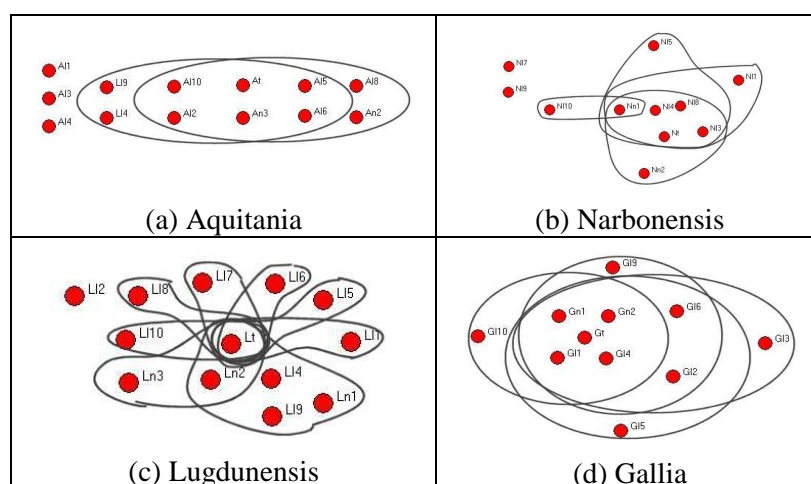
Les interactions sociales jouent un rôle essentiel dans les situations d'apprentissage collaboratif. Lorsqu'il se fait à distance et si le dispositif est en mesure de tracer les interactions homme-homme médiées (CMC : *Computer Mediated Communication*), le cadre des réseaux sociaux est alors adapté à l'étude de la structure communicationnelle et des positions des membres du groupe dans cette structure [PATTISSON 94]. Dans la communauté CSCL, de nombreux chercheurs ont tenté d'appliquer ce cadre d'analyse aux interactions en ligne : [DE LAAT et al. 07 ; MARTINEZ et al. 03]. Dans [NURMELA et al. 99] les auteurs précisent qu'en s'intéressant aux relations entre les individus plutôt qu'aux propriétés des individus, l'analyse des réseaux sociaux est plus à même de révéler des structures et des caractéristiques globales du groupe.

Les auteurs de [AVIV et al. 03] analysent les cliques, et les combinent avec une analyse manuelle des contenus. Ils ont montré que les groupes qui réussissent à maintenir une bonne cohésion, grâce à une participation importante, sont aussi ceux qui réalisent les meilleures phases de « *Critical thinking* ». De même, dans [MARTINEZ et al. 03] les auteurs conjuguent analyses qualitatives de contenus et analyse des réseaux sociaux.

Dans cette nouvelle étude, nous avons entrepris une seconde analyse d'un corpus déjà utilisé dans [REFFAY & CHANIER 03a, 03b] pour mesurer la cohésion d'un groupe. C'est en prenant ces résultats initiaux comme référence, que nous montrons ici comment les nouvelles techniques peuvent améliorer l'analyse de la cohésion d'un groupe. Pour ce faire, nous rappelons les principaux résultats de l'analyse de 2003.

## 2.4. Rappel des principaux résultats de la première analyse

Dans [REFFAY & CHANIER 03a], le seuil d'intensité était fixé à une valeur ( $s=10$ ) justifiée par le nombre de messages demandés dans le scénario pédagogique. Pour ce seuil, les cliques des quatre groupes de base ont été calculées et représentées sur la figure 1. À noter que pour être liés dans le sociogramme, deux individus devaient avoir échangé au moins 10 messages pendant la formation : somme des messages émis par l'un et « lu » par l'autre. Chaque message ne pouvait être « lu » qu'individuellement, mais l'interface permettait à l'utilisateur de marquer un ensemble de messages comme « lu » sans les ouvrir.



**Figure 1.** Cliques de niveau 10 pour les forums de discussion de Simuligne [REFFAY & CHANIER 03a].

Ces représentations (manuelles) permettaient déjà de comparer la cohésion de ces 4 groupes, mais sans différencier les relations de valeurs supérieures ou égales à 10 alors qu'elles peuvent dépasser 200 messages échangés. Dans [REFFAY & CHANIER 03b], les auteurs proposent de représenter les clusters hiérarchiques pour mieux choisir ce seuil. Mais, comme le montre la partie suivante, l'outil KSV nous affranchit de ces contraintes et permet une exploration aisée de toutes les valeurs d'intensité.

## 3. KSV : un outil détourné

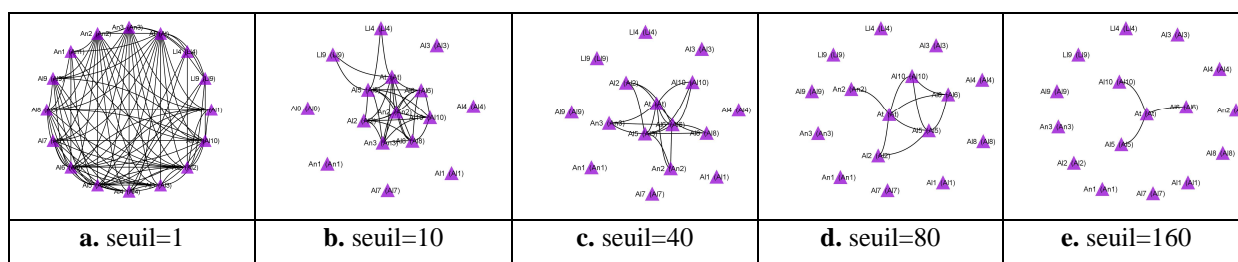
Le "Knowledge Space Visualizer" [KSV], développé par C. Teplovs [TEPLOVS 10], est un outil de visualisation pour faciliter l'analyse des interactions produites dans l'environnement Knowledge Forum (KF) [SCARDAMALIA & BEREITER 03]. Il permet l'exploration des réseaux constitués par les messages (nœuds) et relations de dépendance de deux types : structurelles (répond à, commente, synthétise, etc.) permises par le KF d'une part, et sémantiques calculées automatiquement par la *Latent Semantic Analysis* (LSA) [LANDAUER et al. 98], d'autre part [FUJITA & TEPLOVS 09 ; TEPLOVS & SCARDAMALIA 07].

Dans sa version originelle, KSV visualise un réseau constitué de nœuds et de relations. Les nœuds sont de deux types : contributions dans le forum et auteurs, tandis que les relations peuvent être structurelles (entre auteur et contribution elles sont de type « est l'auteur de », entre 2 contributions « répond à » ou « commente ») ou sémantiques. Parmi les multiples possibilités offertes, l'interface permet à l'utilisateur de définir la période (début-fin) durant laquelle représenter les nœuds (et relations), et éventuellement de la restreindre à un auteur ou aux messages contenant une chaîne de caractères. Les liens sémantiques sont calculés et valués (entre -1 et 1) à l'avance selon la LSA pour toute paire de nœuds du graphe. L'utilisateur, peut alors choisir à l'aide d'un simple curseur : le seuil (ici entre -1 et 1) à partir duquel on peut considérer que 2 contributions sont très proches sémantiquement. Seuls les arcs dont la valeur atteint ce seuil sont alors visibles sur le réseau. Une simulation dynamique utilise l'algorithme de placement dirigé par les forces (attraction - répulsion) [FRUCHTERMAN et al. 91]. Cette simulation rapproche les nœuds liés en éloignant ceux qui ne le sont pas.

Pour conduire la présente étude, KSV a été adapté pour prendre en compte de nouveaux arcs valués entre acteurs plutôt qu'entre contributions. Dans la représentation interne, on définit un arc entre les acteurs A et B si au moins un message a été échangé entre A et B. La valeur de cet arc est le nombre de messages échangés, c'est-à-dire la somme des messages postés par A et lus (ouverts) par B et de ceux postés par B et ouverts par A. Nous avons fait ce choix pour attacher autant d'importance à l'acte de lecture qu'à celui d'écriture mais surtout parce que nous voulions nous baser sur la même définition de relation que celle utilisée dans l'étude précédente [REFFAY & CHANIER 03a, 03b].

#### 4. Nouvelle exploration des réseaux sociaux dans le corpus Simuligne

Dans cette étude, nous nous intéressons plus particulièrement aux échanges dans les forums des 6 espaces (4 groupes de base, Formateurs et Monde) qui ont eu lieu du 3 avril au 31 mai : date de la fermeture du groupe Lugdunensis. La version adaptée de KSV mettant à jour immédiatement la représentation graphique du réseau, nous avons pu facilement analyser les cliques de chacun de ces 6 groupes pour chaque valeur d'intensité (environ 600 analyses). À titre de comparaison, nous n'avions pu en construire (manuellement) qu'une dizaine dans le processus mis en œuvre dans [REFFAY & CHANIER 03a]. La facilité offerte par KSV pour manipuler le seuil d'intensité permet à l'utilisateur de parcourir cette nouvelle dimension à loisir en visualisant les formes que prennent les sociogrammes pour chaque valeur d'intensité. Cette exploration est illustrée sur la figure 2 pour le groupe Aquitania et pour 5 valeurs d'intensité différentes. On retrouve au centre de la figure 2-b, l'une des deux 8-cliques (à droite de la figure 1-a) déjà identifiée dans l'étude de 2003.



**Figure 2.** Les communications dans le groupe Aquitania du 3 avril au 31 mai pour différents seuils d'intensité.

Sur la figure 2-a, il suffit que deux acteurs aient échangé un message (seuil=1) pour que leurs nœuds soient reliés sur le sociogramme. Cette représentation permet de constater que tous les acteurs sont connectés (à au moins 6 autres). En relevant ce seuil à 10 (figure 2-b), on élimine les liens de valeur inférieure à 10. On peut voir la 8-clique (au centre de 2-b) qui est celle du groupe Aquitania avant transfert des 2 apprenants (L14 et L19) de Lugdunensis. Les figures 2-c à 2-e permettent de situer les 8 acteurs les plus communicants entre eux : du centre vers la périphérie, on trouve : At (tuteur) ; A15, A16, A110 ; A12, An2 ; An3, A18.

Grâce à KSV, il ne s'agit plus de choisir un seuil d'intensité pour analyser la structure, mais plutôt de rechercher les formes singulières le long de la dimension intensité.

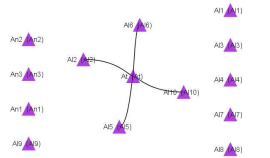
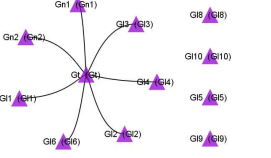
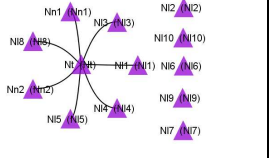
## 5. Résultats

### 5.1. Quelques avancées pour l'analyse des réseaux sociaux

Cette exploration (pour chaque groupe) nous a conduits au model cœur/périphérie de [BORGATTI & EVERETT 99] en considérant que les individus les plus reliés (en nombre de voisins et en intensité) sont au cœur de la structure de communication du groupe, tandis que les autres en sont à la périphérie. Pour un seuil suffisamment élevé ( $c_3+1$ ), nous obtenons généralement une étoile (l'étoile de Freeman [FREEMAN 79]). Nous définissons «  $c_3$  » comme l'intensité la plus grande pour laquelle on trouve au moins une « 3-clique » (sous-ensemble de 3 nœuds tous liés entre eux), qui se manifeste par un triangle dans le graphe. Par conséquent, pour un seuil de «  $c_3+1$  », plus aucun triangle ne peut apparaître. En théorie, de nombreuses formes sont possibles, mais sur nos 6 groupes, nous n'avons trouvé que des étoiles. Pour chaque groupe, l'intérêt était de savoir : Pour quel seuil ( $c_3+1$ ) ? Qui est au centre de l'étoile ? Combien y a-t-il de branches ? Qui en fait partie ?

Nous pensons que ces informations constituent des indications très précises sur la forme de communication des acteurs du cœur du groupe. Ces indications sont données pour les 6 groupes de Simuligne dans les tableaux 1 et 2.

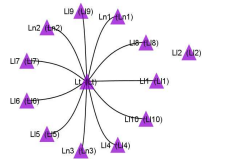
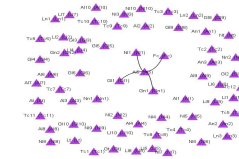
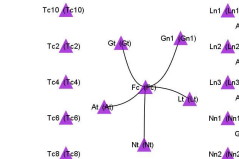
En observant les tableaux 1 et 2 on remarque que parmi les 4 groupes de base (lieux de la simulation globale), et pour l'ensemble des indicateurs, Gallia et Narbonensis sont très semblables tandis que le seuil et le nombre de branches de l'étoile placent Aquitania et Lugdunensis aux valeurs extrêmes.

Indicateurs\groupe	<i>Aquitania</i>	<i>Gallia</i>	<i>Narbonensis</i>
Forme du réseau			
Seuil ( $c+1$ )	129	36	49
Nb. branches ?	4	7	7
Centre ?	Tuteur	Tuteur	Tuteur
Branches ?	4 Apprenants	5 Apprenants, 2 Natifs	5 Apprenants, 2 Natifs

**Tableau 1.** Indications clés au seuil « étoile » pour les groupes Aquitania, Gallia et Narbonensis.

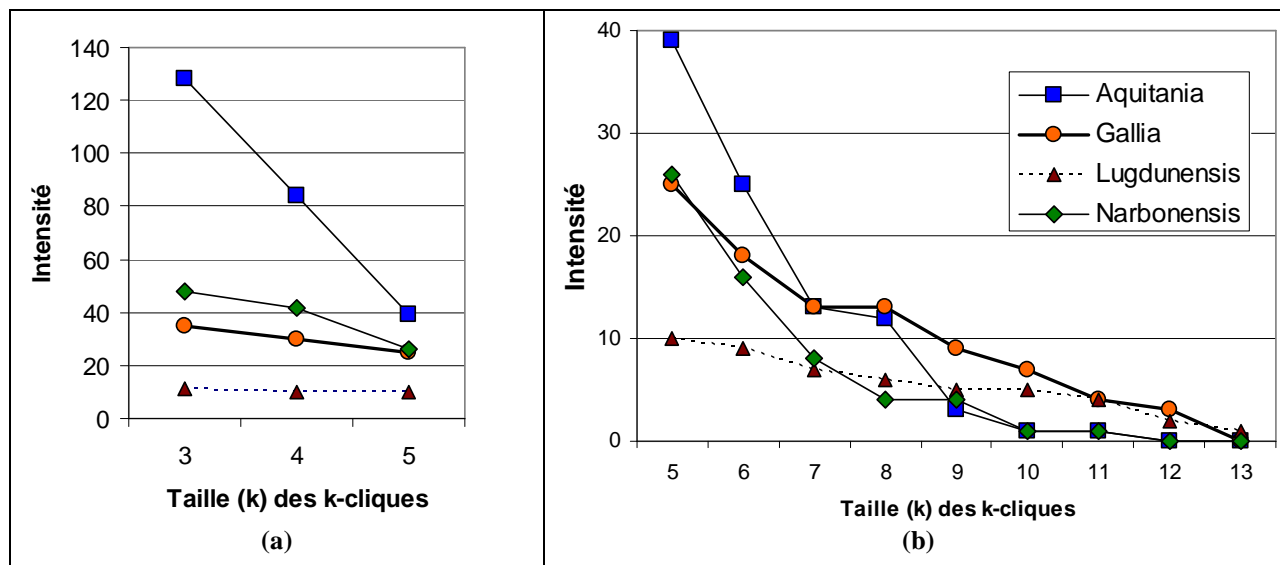


Dans ces 4 groupes, le tuteur est toujours au centre de l'étoile. Aquitania est le seul groupe dans lequel aucun natif n'est une branche de l'étoile. Le centre du Monde (lieu de l'activité interculturelle, où chacun participe en tant que représentant d'un pays, d'une langue et d'une culture) est un apprenant et non un tuteur. Par contre, le coordinateur se trouve bien au centre de l'étoile du groupe Formateurs.

Indicateurs\groupe	<i>Lugdunensis</i>	<i>Monde</i>	<i>Formateurs</i>
Forme de l'étoile			
Seuil ( $c+1$ )	12	77	121
Nb. branches ?	11	4	5
Centre ?	Tuteur	Al5 (Apprenant)	Coordinateur
Branches ?	8 Apprenants, 3 N.	2 App., 1 N, 1 coord.	4 Tuteurs, 1 Natif

**Tableau 2.** Indications clés au seuil « étoile » pour les groupes *Lugdunensis*, *Monde* et *Formateurs*.

En généralisant cette notion de seuil « étoile » atteint pour  $c_3+1$ , nous nous sommes intéressés aux valeurs  $c_k$  (avec  $k \in \mathbb{N}$ , et  $k \geq 3$ ). Les figures 3a et 3b présentent, pour chaque groupe, le niveau d'intensité maximum de ses  $k$ -cliques (sous-ensembles de  $k$  individus tous reliés entre eux pour une intensité donnée). Sur ces graphiques, la taille ( $k$ ) des  $k$ -cliques est représentée en abscisse et l'intensité maximale en ordonnée.



**Figure 3.** Intensités maximales de communication des  $k$ -cliques pour les 4 groupes de base.

Pour obtenir une meilleure précision dans la lecture des intensités pour les cliques de taille  $k > 5$ , le graphique a été scindé (à  $k=5$ ) en 2 parties (a) et (b) avec des échelles d'intensité différentes. Sur la figure 3(a), le premier point de la courbe de Aquitania ( $k=3$ , intensité=128) signifie que la plus grande intensité partagée dans une 3-clique de Aquitania est 128. Les valeurs seuil « étoile » des tableaux 1 et 2 correspondent (à une unité près) aux ordonnées de l'abscisse  $k=3$  : Aquitania : 128, Gallia : 35, Lugdunensis : 11, Narbonensis :

48. Ces courbes représentent une signature très précise de la structure de communication de chaque groupe.

Par construction, ces courbes sont nécessairement décroissantes. En effet, une  $k$ -clique étant composée de «  $(k-1)$ -cliques », ces sous-cliques sont nécessairement de niveau au moins égal à celui de la  $k$ -clique englobante. La forme idéale attendue pour ces courbes dépend du contexte : scénario pédagogique, domaine d'apprentissage. Dans le cas de Simuligne, puisque l'objectif était de faire participer un maximum d'acteurs avec la plus grande intensité possible, on souhaite que, malgré la décroissance naturelle, le niveau (intensité) ne soit pas trop bas ( $\text{intensité} \geq 10$ ) même pour les  $k$ -cliques de grande taille ( $k > 6$ ).

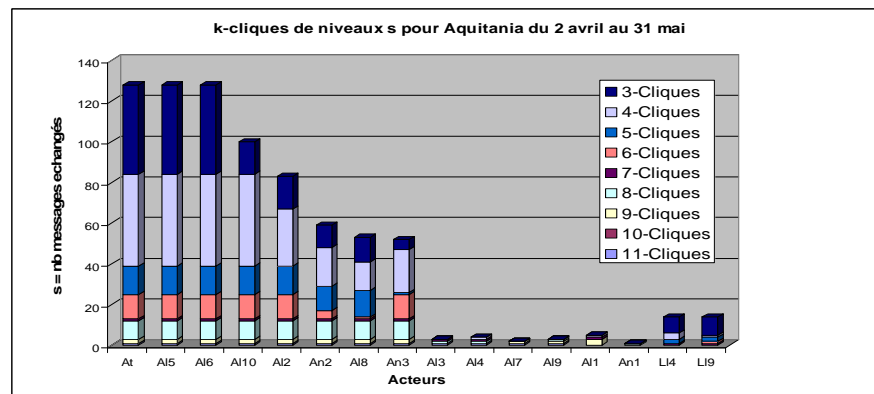
Si l'analyse des indicateurs au seuil « étoile » avait montré la ressemblance entre les groupes Gallia et Narbonensis, on constate que leurs courbes sont très proches pour  $k \leq 6$  mais se séparent nettement en intensité pour les cliques de 7 membres ou plus. Pour  $7 \leq k \leq 11$ , c'est le groupe Gallia qui affiche les plus grandes intensités. On note que les deux groupes qui ont le plus grand seuil « étoile » (Aquitania et Narbonensis : tableau 1) connaissent une décroissance vertigineuse (Fig.3) : une communication intense dans le petit noyau central mais qui diminue très vite au fur et à mesure que le noyau est élargi. À l'inverse, le groupe Gallia dont le seuil « étoile » (tableau 1) n'arrive qu'en troisième position réussit à maintenir une intensité de communication plus élevée dans des cliques de plus grande taille. Enfin, les intensités du groupe Lugdunensis sont si basses que : si dans [REFFAY & CHANIER 03] on avait choisi un seuil de 12 (au lieu de 10) pour l'analyse des cliques, on n'aurait trouvé aucune 3-clique alors qu'on trouve des 8-cliques d'intensité 12 dans Aquitania et Gallia.

Si l'intensité du noyau de Aquitania est la plus élevée (plus du double de celle de Narbonensis, pourtant la plus grande des 3 autres), celle de Gallia domine la deuxième partie du graphe (Fig 3-b) pour  $k \geq 7$ . Les courbes de Aquitania et Narbonensis passent en deçà de celle de Lugdunensis pour  $k > 8$ . Cette observation nous suggère une nouvelle hypothèse :

Une communication trop intense dans un noyau restreint ne risque-t-elle pas de décourager les acteurs périphériques ?

Ainsi, l'analyse des cliques, le long de la dimension intensité montre que (1) les cliques (même petites) de Lugdunensis sont d'une très basse intensité, (2) Aquitania et Narbonensis présentent des cliques qui évoluent de la même façon : un noyau restreint avec une communication très intense et relativement peu de communication dans les cliques plus grandes et (3) Gallia a un noyau qui ne communique pas autant que les deux précédents, mais c'est le groupe qui réussit à maintenir le plus haut niveau d'échange dans les cliques de plus grande tailles.

Si cette exploration permet de mieux comparer les groupes entre eux, elle donne aussi de nouvelles possibilités pour situer chaque acteur dans la structure. En faisant varier le seuil d'intensité du maximum au minimum (1) (ex : en lisant la figure 2 de droite à gauche de 2e à 2a), on peut voir pour quelle intensité, chaque acteur « entre » dans une  $k$ -clique. Le graphique donné figure 4 présente pour chaque acteur (en abscisse), l'intensité (en ordonnée) pour laquelle il entre dans chaque  $k$ -clique. Par exemple, A12 entre dans une 3-clique avec une intensité de 83 messages échangés avec chacun des deux autres membres puis dans une 4-clique avec une intensité de 67. Les 8 acteurs (à gauche) entrent dans leur 8-clique avec une intensité de 12.



**Figure 4.** Graphe des intensités auxquelles les membres de Aquitania entrent dans les k-cliques.

Le graphe donné en figure 4 montre comment le cœur (au sens de [BORGATTI & EVERETT 99]) s'étend (k allant de 3 à 8) au fur et à mesure que l'intensité diminue (de 128 à 12). Si l'on pouvait générer cette représentation automatiquement pendant la formation, elle permettrait, par exemple au tuteur, de détecter aisément les acteurs qui échangent le plus avec le plus grand nombre et surtout, ceux qui échangent le moins. Les premiers pouvant servir d'appui pour aider les seconds à plus interagir avec le groupe.

### 5.2. De nouvelles perspectives pour l'outil KSV

Pour la première fois, KSV était adapté pour s'appliquer à des données non issues de l'environnement Knowledge Forum. Ce travail d'adaptation a permis de générer une version de KSV qui soit indépendante du Knowledge Forum travaillant sur le format ouvert [GraphML]. Cette nouvelle version est libre et gratuite (<http://code.google.com/p/ksv>).

Fort de cette nouvelle forme de réseau social (plutôt que graphe des contributions), KSV peut maintenant envisager de conjuguer analyse sémantique et analyse des réseaux sociaux. Pour tenter d'unifier les deux types de nœuds (contribution vs acteur), on peut regrouper l'ensemble des contributions d'un acteur pour représenter (sémantiquement) un acteur par ce qu'il a produit/écrit. On peut aussi envisager d'y ajouter un lien structurel représentant les contributions qu'il a ouvertes (lues).

Ce travail d'adaptation a montré la malléabilité de KSV. Initialement conçu pour visualiser le réseau des contributions dans le *Knowledge Forum*, il a pu s'en détacher et être enrichi en moins d'un mois pour traiter de nouvelles données (ex : nb messages) et réviser certaines hypothèses (ex : lecture des messages) et ainsi offrir toute l'ingéniosité de son interface au service de l'analyse des réseaux sociaux.

## 6. Conclusion

Cette communication est le fruit de la rencontre entre le corpus Simuligne et l'outil KSV. Elle illustre l'intérêt de l'échange d'outils et de corpus pour que des perspectives différentes puissent se nourrir l'une l'autre. Le corpus, transmis dans un format explicite (XML), a été réutilisé par un chercheur extérieur à la formation Simuligne. KSV a montré sa malléabilité et son efficacité en permettant l'exploration des sociogrammes issus des forums de Simuligne, pour toutes les valeurs d'intensité. L'exploration de cette dimension offre plus de

puissance à l'analyse et de nouvelles possibilités. Plutôt que de mesurer la cohésion des groupes pour un seuil fixé et d'y analyser les cliques, on peut désormais, pour chaque groupe, rechercher les seuils au-delà desquels certains phénomènes intéressants se produisent, pour ensuite comparer la valeur de ces seuils entre les différents groupes. C'est ce principe que nous avons appliqué pour repérer le seuil « étoile » de chaque groupe. Puisque notre mesure de cohésion est basée sur la taille et l'intensité des cliques, nous avons proposé ici de comparer les groupes selon la courbe d'intensité des cliques en fonction de leur taille. Ce nouvel indicateur donne bien plus d'épaisseur à l'analyse et nous a permis de mieux caractériser les comportements de chacun des groupes. Enfin, nous avons généré un graphe représentant pour chaque acteur l'intensité de son entrée dans chaque k-clique.

Enfin, suite à ce travail, la communauté dispose d'une version (libre) de KSV qui peut opérer directement sur des données en format ouvert (standard GraphML).

## Remerciements

Les auteurs tiennent à remercier l'École Nationale Supérieure de Cachan et le laboratoire STEF pour avoir pris en charge l'accueil de C. Teplovs en septembre et octobre 2010.

## Bibliographie

- [AVIV et al. 03] Aviv, R., Erlich, Z., Ravid, G., & Geva, A. "Network analysis of knowledge construction in asynchronous learning networks". *Journal of Asynchronous Learning Networks*, 7(3), 1–23. 2003.
- [BORGATTI & EVERETT 99] Borgatti, S. P., Everett, M. G. "Models of core/periphery structures". *Social Networks*, 21, 375–395. 1999.
- [BORGATTI et al. 02] Borgatti, S. P., Everett, M. G., & Freeman, L. C. *UCINET 6 For Windows: Software for Social Network Analysis*. Harvard, MA, 2002.
- [DEGENNE & FORSÉ 94] Degenne, A. & Forsé, M. *Les réseaux Sociaux*. Collection U, série « Sociologie », dirigé par H. Mendras, Armand Colin, 1994.
- [DE LAAT et al. 07] de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis". *International Journal of CSCL*, 2(1), 87–103. 2007.
- [FREEMAN 79] Freeman, L. C. "Centrality in social networks conceptual clarification". *Social Networks*, 1(3), 215–239. 1979.
- [FRUCHTERMAN et al. 91] Fruchterman, T. M. J., Reingold, E. M. Graph Drawing by Force-Directed Placement. *Software – Practice & Experience* (Wiley) 21 (11): 1129–1164. 1991. doi:10.1002/spe.4380211102
- [FUJITA & TEPLOVS 09] Fujita, N., & Teplovs, C. "Automating the Analysis of Collaborative Discourse: Identifying Idea Clusters". In A. Dimitracopoulou, C. O'Malley, D. Suthers & P. Reimann (Eds.), *Computer Supported Collaborative Learning Practices: CSCL 09 Community Events Proceedings*, 162–164. International Society of the Learning Sciences. 2009.
- [FURSTENBERG et al. 01] Furstenberg, G., Levet, S. & Maillet, K. "Giving a virtual voice to the silent language of culture: the cultura project". *LLTJ*, (5), 1, 55–102. January 2001.
- [LANDAUER et al. 98] Landauer, T., Foltz, P. W., Laham, D. "Introduction to Latent Semantic Analysis". *Discourse Processes* 25: 259–284. 1998. doi:10.1080/01638539809545028

- [MARTINEZ et al. 03] Martínez, A., Dimitriadis, Y., Rubia, B., Gomez, E., & de la Fuente, P. "Combining qualitative evaluation and social network analysis for the study of classroom social interactions". *Computers & Education*, 41(4), 353–368. 2003.
- [NURMELA et al. 99] Nurmela, K., Lehtinen, E., & Palonen, T. "Evaluating CSCL log files by social network analysis". In C. M. Hoadley, J. Roschelle, (Eds.), *Proceedings of the Computer Support for Collaborative Learning 1999 Conference*. 434–444. 1999.
- [PATTISSON 94] Pattison, P. E. "Social cognition in context: Some applications of social network analysis". In S. Wasserman & J. Galaskiewicz (Eds.), *Advances in social network analysis in the social and behavioral sciences*, 79-109. Newbury Park, CA: Sage. 1994.
- [REFFAY & CHANIER 03a] Reffay, C., & Chanier, T. « Mesurer la cohésion d'un groupe d'apprentissage en formation à distance ». *Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH'2003)*, Strasbourg, France, pages 367–378, Avril 2003.
- [REFFAY & CHANIER 03b] Reffay, C., & Chanier, T. "How social network analysis can help to measure cohesion in collaborative distance-learning". In B. Wason, S. Ludvigson & U. Hoppe (Eds.), *Designing for change in networked learning. Proceedings of the international conference on Computer Supported Collaborative Learning 2003* 343–352.
- [SCARDAMALIA & BEREITER 03] Scardamalia, M. & Bereiter, C. "Learning to work creatively with knowledge". In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merriënboer (Eds.), *Unravelling basic components and dimensions of powerful learning environments. EARLI Advances in Learning and Instruction Series*. 2003. <http://ikit.org/fulltext/inresslearning.pdf>
- [SCOTT 00] SCOTT, J., *Social Network Analysis: a handbook*, 2<sup>ed.</sup>, SAGE, London, 2000.
- [TEPLOVS 10] Teplovs, C. Visualization of knowledge spaces to enable concurrent, embedded and transformative input to knowledge building processes. Thèse de l'University of Toronto, 2010.
- [TEPLOVS & SCARDAMALIA 07] Teplovs, C., & Scardamalia, M. "Visualizations for knowledge building assessment". Paper presented at the *AgileViz workshop*, CSCL 2007.
- [YAICHE 96] Yaiche, F. *Les simulations globales, mode d'emploi*. Hachette FLE : Paris, 1996.

### **Références sur le WEB.**

Tous les liens présentés dans cette liste ont été visités en mars 2011

[GraphML] : <http://graphml.graphdrawing.org/> : The GraphML file format

[GraphViz] <http://www.graphviz.org/> : Graph visualization software

[KSV] <http://chris.ikit.org/ksv/> : Knowledge Space Visualizer

[Mulce] <http://mulce.org> : projet ANR (2007-2010) coordonné par T. Chanier

[SNAtool] [http://en.wikipedia.org/wiki/Social\\_network\\_analysis\\_software](http://en.wikipedia.org/wiki/Social_network_analysis_software)

[Ucinet 6] <http://www.analytictech.com/ucinet/versions.htm> : UCINET version 6.